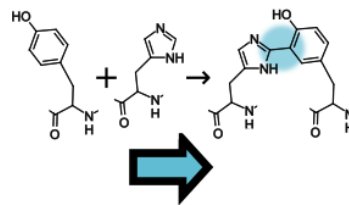


Enzyme Biocuration and Reaction Prediction using Machine Learning

| | |
|---------------------|--|
| Supervisor | Ass.-Prof. Mitja M. Zdouc, Division of Pharmacognosy |
| Topic | Biocuration, data science, machine learning |
| Duration | 6 Months (full-time), computational work only |
| Requirements | Bachelor in Pharmacy, Informatics, or similar |
| Language | English/German |

Description

Many organisms produce **natural products** (NPs), small molecules with antimicrobial and other specialized **biological activities**. The biosynthesis of NPs is orchestrated by one or more so-called **tailoring enzymes**, which are often responsible for enhanced activity and bioavailability. Despite their **importance** for **bioinformatics** and **drug discovery**, information on tailoring enzymes is habitually deposited in **narrative articles**, hampering computational access. To make tailoring enzyme data **Findable, Accessible, Interoperable, and Reusable** (FAIR), we have developed the **MITE database** [1]. Freely available at <https://mite.bioinformatics.nl/>, this knowledge base of **reaction- and substrate-specificity** can be used for **annotation**, as a dataset for **machine learning**, and as a parts catalog for **synthetic biology**.



In this project, you will expand the coverage of the MITE database in preparation of the MITE version 2.0 publication, planned for 2027. You will perform **biocuration** [2], querying the literature for publications on tailoring enzymes, critically evaluating the presented experimental evidence, summarizing key information in form of controlled vocabularies, and cross-linking to related databases, thereby contributing to **linked open data**. You will analyze the collected data and apply **machine learning**, for instance to perform **reaction predictions** of **unknown tailoring enzymes**. This work will serve as a foundation for better prioritization algorithms that speed up natural product discovery.

Learning outcomes

At the end of this Master thesis, you will be able to:

- **Discuss** the enzymatic logic of NP biosynthesis and **distinguish** biosynthetic pathways.
- **Apply** biocuration methods by critically evaluating literature and **synthesize** knowledge using controlled vocabularies, cross-referencing to related data sources.
- **Apply** machine learning algorithms in predicting reactions of unknown tailoring enzymes.
- **Present** findings in word and text and **understand** the manuscript writing process.

How to apply

Please contact Ass.-Prof. Mitja Zdouc via E-Mail (mitja.zdouc@univie.ac.at), indicating the progress of your studies, your preferred starting date, and if you fulfill the requirement criteria. Please include a brief statement about your research interest, why you would like to work on this project, any prior knowledge related to the topic, as well as preferred working style and supervision expectations. Applicants will be invited for a brief interview where the project will be discussed in detail.

References:

- 1 Rutz et al. 2025 Nucl Acids Res DOI: [10.1093/nar/gkaf969](https://doi.org/10.1093/nar/gkaf969)
- 2 <https://en.wikipedia.org/wiki/Biocuration>